‘hagis’, an R Package Resource for Pathotype Analysis of *Phytophthora sojae* Populations Causing Stem and Root Rot of Soybean

Austin G. McCoy\*,1,✉, Zachary Noel\*\*,1, Adam H. Sparks\*\*,2, and Martin Chilvers1,3

19 June, 2019

*Phytophthora sojae* is a significant pathogen of soybean worldwide. *P. sojae* pathotype surveys are regularly conducted to monitor resistance gene efficacy and determine if new resistance genes need to be deployed. Valuable measurements for pathotype analysis include the distribution of susceptible reactions, pathotype complexity, pathotype frequency, and diversity indices for pathotype distributions. Previously the Habgood-Gilmour Spreadsheet (HaGiS), written in Microsoft Excel, was used to obtain pathotype data. However, the growing popularity of R in plant pathology and desire for reproducible research made HaGiS a prime candidate for conversion into a more replicable form. Here we report on the creation and use of an R package, `hagis’, that can be used to produce all outputs from the HaGiS Microsoft Excel sheet for *P. sojae* pathotype studies.

\* These authors contributed equally to this work.

1 Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA  
2 University of Southern Queensland Centre for Crop Health, Toowoomba, Qld 4350, AUS  
3 Program in Ecology, Evolutionary Biology, and Behavior, Michigan State University, East Lansing, MI 48824, USA

✉ Correspondence: [Austin G. McCoy <[mccoyaus@msu.edu](mailto:mccoyaus@msu.edu)>](mailto:mccoyaus@msu.edu)

# R Package Announcement

Uniform and healthy stand establishment is essential to maximizing soybean yield. Oomycetes such as *Phytophthora sojae* constitute a significant threat to soybean (*Glycine max*) stand establishment and yield. *P. sojae* has been managed primarily via deployment of single resistance genes in commercial soybean cultivars, which interact with *P. sojae* *Avr* gene products to confer resistance (Anderson et al. 2015). Genetic resistance to *P. sojae* is the most economical form of control for *P. sojae* as it confers season-long protection to non-compatible pathotypes (Dorrance et al. 2016). However, *P. sojae* pathotype surveys need to be regularly conducted to determine shifts in pathotypes over time and provide recommendations for effective resistance genes. Although statewide pathotype surveys have been conducted for the past 60 years, there has been no significant advance in pathotype analysis since the development of the Habgood-Gilmour Spreadsheet (HaGiS), written in Microsoft Excel, in 1999 (Herrmann et al. 1999; Kaufmann et al. 1958).

*Phytophthora sojae* pathotype studies study the efficacy of soybean resistance genes in relation to a, or multiple, states *P. sojae* population. In doing so, very large sets of virulence data are generated for potentially hundreds of isolates (Dorrance et al. 2016). Using such large datasets within the HaGiS Excel based program can be cumbersome and time intensive to transfer the data into and analyze. R offers the ability to work with large data sets, such as pathotype survey virulence data, in an easy and efficient manner without the additional steps of data entry that the HaGiS Excel program requires.

The R language provides an alternative to Microsoft Excel based data analysis and has become widely used in plant pathology studies due to its open-source framework and amenability to conduct reproducible research (R Core Team 2019; Sparks et al. 2008; Bergna et al. 2018; Wallace et al. 2018). Using an R package for analyzing pathotype survey data can replicate all analyses provided by Excel-based programs. It allows users to create reproducible research and more detailed visualizations as well as allowing the plant pathology community to actively contribute to and build upon this code for future studies. For instance, McCoy and Noel produced R scripts to conduct these analyses originally performed with HaGiS, which were used to create the ‘hagis’ R package (McCoy and Noel 2018; McCoy et al. 2019).

‘hagis’ can be installed from the Comprehensive R Archive Network (CRAN) using the R command, install.packages(“hagis”). For ease of use, the package uses a single argument format, which works in all ‘hagis’ functions. Users provide their own data in the form of a spreadsheet, CSV or text file, specifying the proper fields for analysis when using ‘hagis’. The results are returned as tables or figures as appropriate. Functions are provided to calculate pathotype complexity and summarize the distribution of reactions for each gene tested. Diversity indices that are calculated include Simple, Shannon, Simpson, Gleason and Evenness indices for the pathotype data set. Outputs from these analyses are given in publication ready graphics or tables and can be further modified by the user.

The R language offers many advantages to Excel based data analysis such as reproducibility, user customization and the ability to work efficiently with large data sets such as those produced through *P. sojae* pathotype surveys. Significantly, this R package provides the first development of *P. sojae* pathotype analysis in 20 years. Likewise, openly available R code for these analyses will provide the plant pathology community with an alternative to Excel based data analysis for improved reproducibility. ‘hagis’ was produced for *P. sojae* pathotype surveys, but can be used in other gene-for-gene systems to determine effective resistance genes in management. More information and instructions on how to use `hagis’ can be found at: <https://openplantpathology.github.io/hagis/>.

# Acknowledgements

Funding for this work was provided by: Michigan Soy, Project GREEEN, *Dorrance Grant information Here* and GRDC Project DAQ00186 - Improving Grower Surveillance, Management Epidemiology Knowledge And Tools To Manage Crop Disease

# References

Anderson, R. G., Deb, D., Fedkenheuer, K., and McDowell, J. M. 2015. Recent progress in RXLR effector research. Molecular Plant-Microbe Interactions. 28:1063–1072

Bergna, A., Cernava, T., Rändler, M., Grosch, R., Zachow, C., and Berg, G. 2018. Tomato seeds preferably transmit plant beneficial endophytes. Phytobiomes Journal. 2:183–193

Dorrance, A., Kurle, J., Robertson, A., Bradley, C., Giesler, L., Wise, K., and Concibido, V. 2016. Pathotype diversity of *Phytophthora sojae* in eleven states in the united states. Plant disease. 100:1429–1437

Herrmann, A., Lower, C., and Schachtel, G. 1999. A new tool for entry and analysis of virulence data for plant pathogens [correspondence]. Plant Pathology (United Kingdom).

Kaufmann, M. J., Gerdemann, J., and others. 1958. Root and stem rot of soybean caused by *Phytophthora sojae* n. Sp. Phytopathology. 48

McCoy, A., and Noel, Z. 2018. AGmccoy/Phytopthora-sojae-Pathotype-analysis: Beta-release of Phytophthora sojae pathotype analysis code.

McCoy, A., Noel, Z., Sparks, A. H., and Chilvers, M. 2019. *hagis: Analysis of plant pathogen pathotype complexities, distributions and diversity*.

R Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Sparks, A. H., Esker, P., Bates, M., Dall’Acqua, W., Guo, Z., Segovia, V., Silwal, S., Tolos, S., and Garrett, K. A. 2008. Ecology and epidemiology in R: Disease progress over time.

Wallace, J. G., Kremling, K. A., Kovar, L. L., and Buckler, E. S. 2018. Quantitative genetics of the maize leaf microbiome. Phytobiomes Journal. 2:208–224